SANS

Analyst Program

Whitepaper

Google's Chronicle Security Operations: Why Doesn't My SIEM Do That?

Written by **Jake Williams**

November 2022

Chronicle

©2022 SANS™ Institute

Introduction

SANS conducted a review of Chronicle, Google's cloud-native security operations suite, with a focus on evaluating its SIEM features and usability from a practitioner perspective. As readers will learn from this review, the product has a significant number of capabilities and was obviously designed to address shortcomings inherent in many SIEM platforms. The interface was easy to navigate and makes operating through traditional analyst workflows seamless. After reading this product review, we believe you'll want to give Chronicle a look for your security operations team. At a minimum, we think you'll ask, "Why doesn't my SIEM do that?" on more than one occasion.

History of the SIEM

In the beginning, there was no focus on security—after all, logs were just for troubleshooting bugs and misconfigurations. Then there was local log review, followed by centralized logs, and finally advanced correlation with SIEM. The SIEM promised to support more advanced alerting by correlating logs from multiple sources, allowing them to be used to generate alerts or even automagically eliminate a false positive detection. Additionally, SIEM allowed organizations to support threat hunting and more comprehensive security investigations.

Although the SIEM was certainly a leap, most organizations found they couldn't realize the full promise of SIEM technology. Sometimes this was due to limits in the number of logs that could be ingested or the amount of retention for those log sources (both a storage limitation). In other cases, organizations found they couldn't index all the right data needed for correlations, usually due to memory constraints. Still others discovered they couldn't enable as many correlation rules as they wanted, primarily due to processing limitations.

When an organization undersized its SIEM's storage, memory, or processing power, it typically needed to wait another budgeting cycle for additional capital expenditure (CAPEX) funds to rectify the situation. Cloud-native SIEMs emerged to address the CAPEX problem and challenges with scale/elasticity. Now organizations could largely pay by the volume of logs ingested, allowing the cloud-native SIEM provider to deal with the back-end issues of hardware. Organizations could scale up or down with relative ease, so long as they paid the bill for the volume of data ingested.

But even cloud-native SIEMs failed to address other core problems with SIEM deployments. Hardware aside, most organizations struggle to determine how to operationalize the logs they have. Sure, the SIEM provides visibility into logs, but the organization needs to know what to look for to create detections. There's a significant gap between the detections in more basic and advanced organizations. Although the SIEM is a very capable tool, it is just that—a tool. A carpenter does amazing things with woodworking tools that a novice cannot hope to replicate. So, too, is it with most SIEMs (until now, that is). And once a security event is detected, how is it handled? The SIEM must effectively support the responders who investigate and remediate security events. Read on to learn how Chronicle is changing the game in these regards.

SIEMs in general were supposed to enable three main goals:

- Increasing visibility
- Enabling detection of security events
- Supporting the response of discovered security issues

Although those goals haven't changed, it's undeniable that traditional SIEMs have failed to address them. Addressing these shortcomings is core to the Chronicle mission.

Chronicle Cloud-Native SIEM

Google's cloud-native SIEM Chronicle is designed from the ground up to address shortcomings found in other SIEMs. As the Chronicle team shared, they don't want this review to be a bake-off of features between Chronicle and other SIEMs. This will be impossible in some areas, because it's hard to understand why features built into Chronicle are powerful without understanding the limitations of traditional platforms.

It was clear during our review that Chronicle is about more than just adding some features to existing SIEM platforms. It's about changing the paradigms around how security investigations are performed. Just as an artist creates their workflow based on the brushes, paints, and canvases (e.g., tools) available to them, so too does the security analyst. Much of the workflow we see in today's security investigations and threat hunting has been driven by available tools and features they support. Chronicle's design makes it appear that they intended to design the tool around the *ideal* analyst workflow.

The entire design of Chronicle SIEM focuses on customer outcomes. There are four pillars of security that Chronicle addresses:

- Provide complete visibility into the security environment.
- Enrich data in the SIEM with Google's threat intelligence and external sources, enabling security analysts to rapidly operationalize it.
- Apply modern threat detection to data ingested into the SIEM, without relying on customers to have dedicated security engineering resources on staff.
- Facilitate seamless response to accelerate the investigation by integrating with SOAR platforms (including Chronicle SOAR, formerly Siemplify).

Search Speed

One of the core problems with SIEMs has traditionally been search speed. The speed of searching data in a SIEM is directly correlated to the amount of data that was indexed and how many of those indexes could stay in RAM at any time. Most complex queries relied on indexes from more tables in memory. Search speed isn't just a convenience factor for analysts. In many cases, it changes the way organizations pursue investigations.

Most analysts won't wait more than a few minutes for a search to complete before pivoting to some other analysis activity. Performing the search and analysis of the returned data asynchronously (after the search returns, perhaps hours later) involves high context switching costs. This, in turn, impacts the quality of the analysis including analyst confidence and morale. Slow searching also forces analysts to change the types of queries they run to return data more quickly. Finally, slow search return times necessarily impact the scope of searches performed by analysts. Analysts often limit the duration of their queries to receive some data quickly for analysis now, rather than waiting potentially hours for a slow search to show them the same data across the entire retention period. This has obvious negative implications on the quality of investigations performed.

Given Google's dominance in the search engine market, it's not surprising to learn that Chronicle is highly responsive with search results. In fact, it's the most responsive SIEM we've ever seen. In our tests, there wasn't a query we ran that took more than a few seconds to respond (see Figure 1). In fact, the UI felt so responsive in returning results that we believed it was likely that our Internet speed was the limiting factor in returning results—quite a significant departure from our expectations. Chronicle advertises sub-second search returns, even across datasets that are petabytes in size.

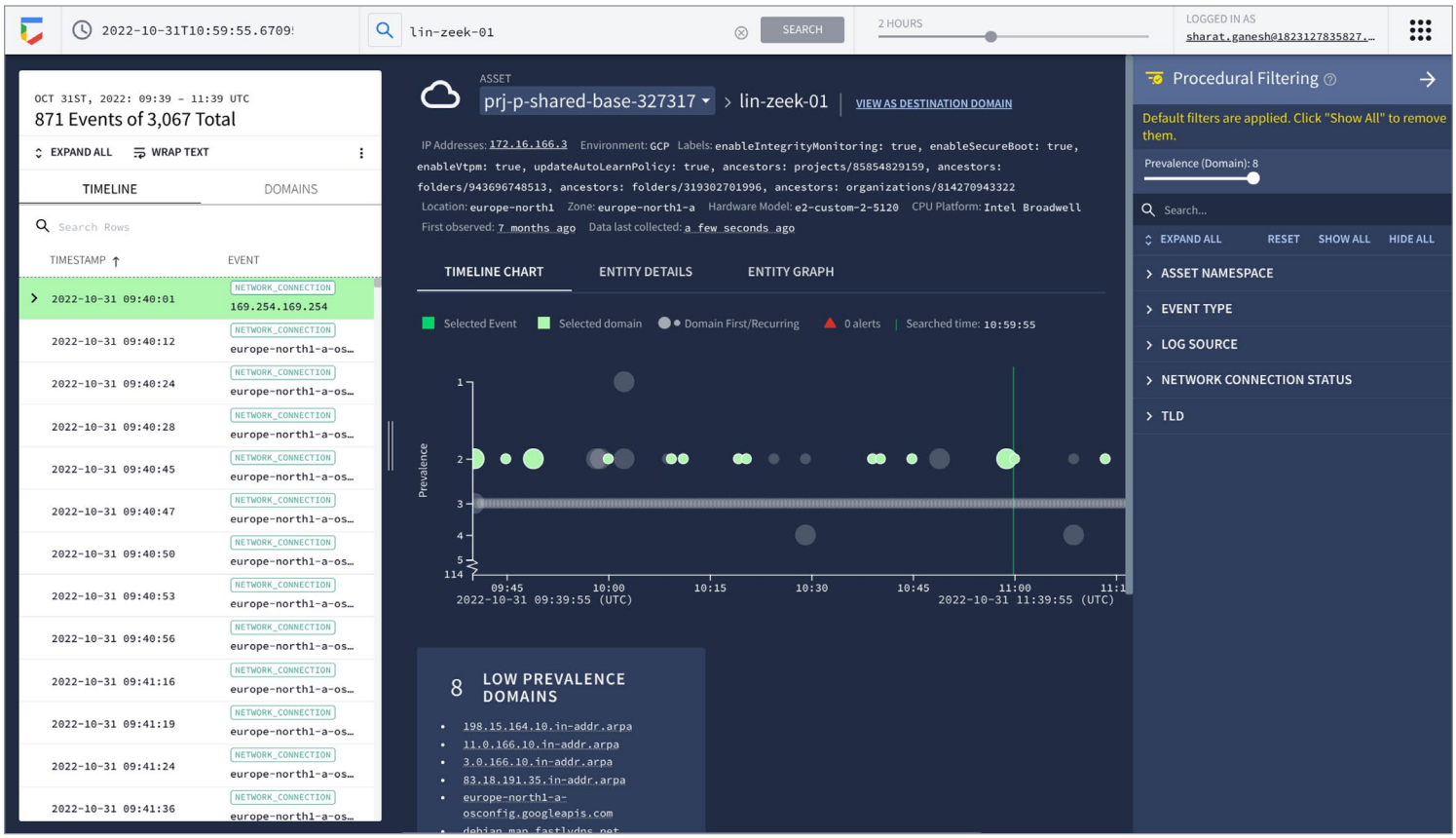

Figure 1. Chronicle SIEM UI with a Query Response on an Asset Search

Hot and Cold Storage

Another common challenge with some SIEM deployments is the idea of hot and cold storage. Because indexing data is expensive in terms of processing and because the index must fit in RAM to achieve high search speeds, many SIEM products move data out of hot storage into cold storage to achieve the required retention periods. For instance, the SIEM may keep 30 days online in hot storage with full indexes and then move data to much slower cold storage to achieve a total retention period of 180 days. This means that searching the first 30 days of data is often fast, but searching beyond that takes exponentially longer. If the analyst is aware of the hot and cold storage configuration, it will necessarily change the way they perform investigations (preferring to limit queries to data in hot storage). If they are unaware of the configuration, things are worse because analysts don't get responses to their queries quickly enough to facilitate complete investigations.

Note: The timeframes for hot and cold storage are just examples and depend on the configuration and hardware constraints of the SIEM deployment.

Chronicle has no notion of hot and cold storage—all data is available at the same blistering speeds. By default, all data ingested into Chronicle is available for one year, though this timeframe can be shortened or lengthened to meet customer demand. This is far more retention than most organizations have with their SIEMs today. Most organizations we consult with have three to six months of storage for their critical logs.

The ability to issue a search and get a positive (or negative) result in seconds across an entire year of data is a game changer for investigations. As threat actors continue to advance their techniques, one of the most powerful tools an analyst has in their arsenal is the knowledge of "have we seen this before?" Chronicle allows analysts to answer this question in seconds across an entire year of data instead of waiting an hour or more to query through a month of the same data.

Data Enrichment

A major feature that helps differentiate SIEM platforms has traditionally been the ability to enrich data in the SIEM and the quality of that enrichment. Many SIEMs limit the fields where logs can be automatically enriched with cyber threat intelligence (CTI). In most cases, these are specific indexed fields where the data is ingested into a particular schema. For instance, consider an IP address. Although an IP address is trivial to identify in an Apache web server log file (it's in a dedicated column, separated by whitespace), most SIEMs will not identify an IP address buried in the description of an application event log entry with no schema.

To be clear, finding a valid IP address is as easy as performing a regular expression (regex) search in a log entry. The problem isn't that that discovery is algorithmically difficult. It's that the processing power required to do this with every log entry is computationally expensive. Unfortunately, this is a cost that most cloud SIEM providers just don't expend for their customers. That's not to mention that IP addresses are easy to find. Other data types are far less so. But this is another area where Chronicle separates itself from its competitors. Every piece of data ingested is parsed for fields to enrich.

Unfortunately, it is extremely difficult for an organization consuming CTI to grade its quality, other than through anecdotal cases where the CTI (or lack thereof) contributed to the success or failure of an investigation. The quality of data enrichment is dependent on a number of factors, but primarily relies on the quality of the SIEM provider's CTI, and that quality is largely driven by the breadth of data an analyst has access to. It's hard to argue against the idea that Google has access to quantities of threat data that would make most analysts drool with jealousy (which certainly contributes to the quality of analysts Google can both hire and retain). Unparalleled visibility into threat data and analysts performing at the very apex of the industry are what feed the CTI Google brings to Chronicle, ensuring your security team's success.

Nowhere is the difference between Chronicle and other SIEMs so clear as in the case of data enrichment. Instead of only enriching data on some fields for alarms, Chronicle enriches data on ingest and then stores that enriched data, making it searchable by analysts. This is an absolute game changer for security analysts.

For an example, let's consider an IP address again. Data enrichment may include:

- Netblock or ASN of the IP address
- Malware samples beaconing to the IP address
- Passive DNS history showing domains pointing to that IP address
- Threat intelligence feeds where the IP address has been listed
- Other IP addresses in the same netblock that have also been listed in threat intelligence feeds

Let's consider passive DNS results. By providing this data to the analyst in a searchable field, it is trivial for the analyst to pivot from a malware detection beaconing to an IP address. The analyst is immediately presented with domains that resolved to the IP address in question. Armed with this information, the analyst pivots in the SIEM to look for communications to other IP addresses that also resolve to domains of interest. This is just one example of advanced security investigation uniquely enabled by Chronicle, thanks to it making enriched data searchable (instead of just limiting searches to source data like traditional SIEMs).

Another example of a place where enrichment is critical is integration with identity platforms. Consider the desire to search for events involving any identity that involves a particular attribute, such as work role (see Figure 2). In addition to using this for understanding how threat actors are targeting victims within the organization, it is a game changer for insider threat investigations. This is doubly true for insider threat investigations where there are multiple investigation targets. Alerting rules can be configured trivially for all users with a particular attribute added to the IDP, possibly by creating a user-specified tag in the IDP or adding the user to a particular group.

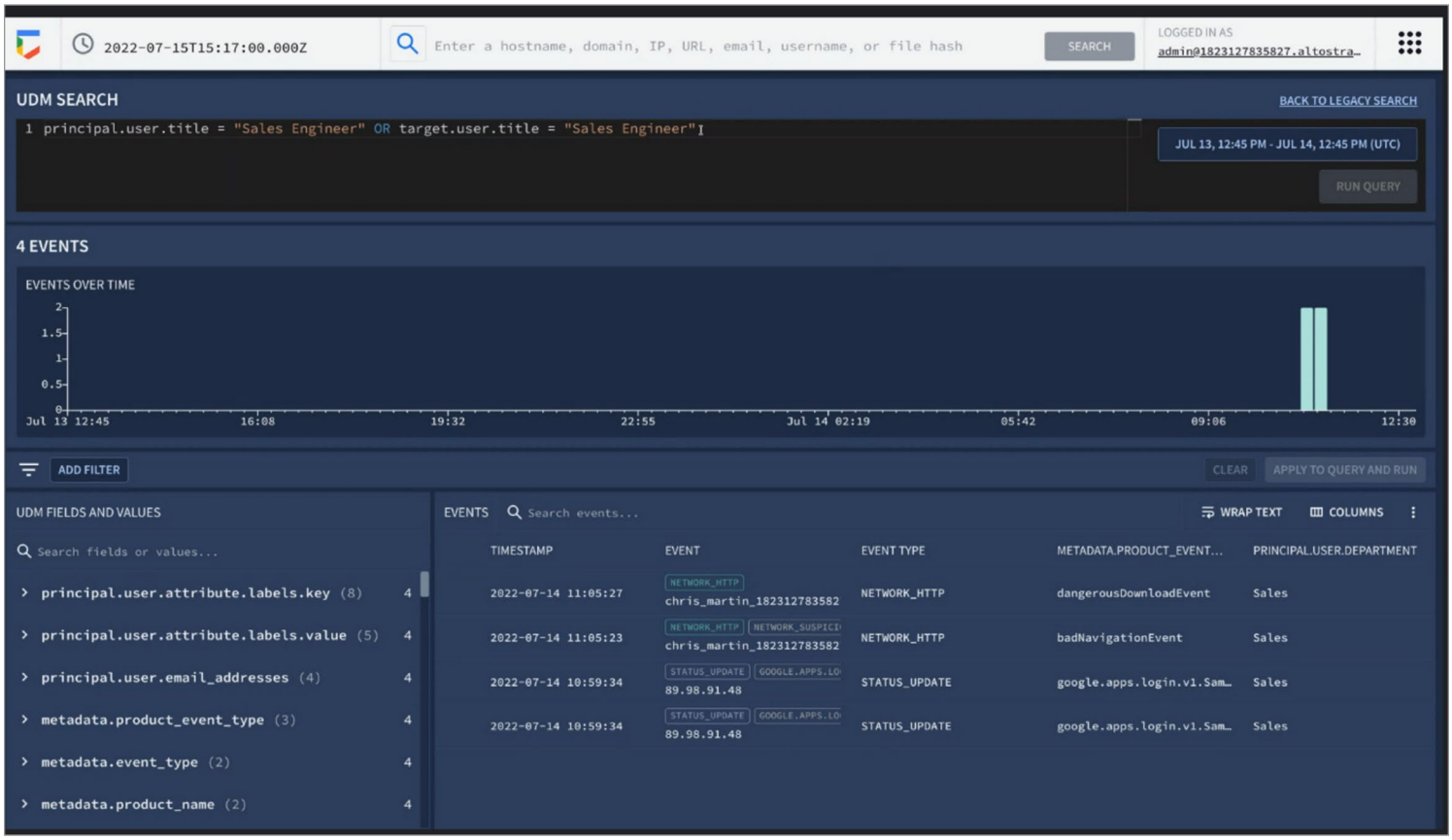

Figure 2. Querying on Identity Roles Not Included in Source Log Data

Hash Aliasing

Hash aliasing is another critical feature of Chronicle. Some CTI feeds support only a single algorithm for file hashes, such as MD5. But suppose that your endpoint detection and response (EDR) logs processes created in SHA1? And of course, your information sharing and analysis center (ISAC) shares indicators of compromise using SHA256. This is a real problem for analysts because there is no way to translate between hash types without the source data (which is most often missing).

Chronicle supports taking a known hash, such as MD5, and translating it to a SHA1, SHA256, or other hash as necessary to perform detections. This is the automation of a manual process analysts often use querying online CTI platforms, hoping to find a Rosetta Stone of sorts. But in addition to being manual, the process is also limited by the visibility the platform has. Chronicle owns the number-one platform used for this activity (VirusTotal) and uses this, in addition to other data sources, to provide hash aliasing for a significant percentage of overall data in our tests.

Taking the Full Data Stream

Because Chronicle can operate effectively against data sets that would grind other SIEM platforms to a halt, it ingests full telemetry from supported EDR platforms. Most other SIEM platforms can only ingest alarms from EDRs, significantly limiting the data available for investigations. Most EDR platforms only store telemetry for 14 or 30 days in default configurations, but as previously mentioned, Chronicle by default stores all data for a year.

Most EDR platforms limit retention because they can't search through it efficiently. It is not unusual for an EDR search across just a week or two of data to take 10 minutes or more to return.

The ability to search across larger periods of data is critical for analysts in many cases. Consider, for instance, the SolarWinds incident. Identification of the intrusion took almost nine months, meaning that most organizations were left without the ability to look at full device activity. Most organizations that had logs going back that far at all only possessed data from domain controllers and, even then, that was largely limited to login event information. But with full telemetry (and the ability to query it), investigations change from "what little data we have is inconsistent with the described event" to "we can affirmatively state from our data that the event did not occur."

That every hash observed in telemetry throughout an entire year is searchable in seconds changes the way investigations are performed. It also changes the game for ingesting threat intelligence. If your organization receives information about a suspicious file hash, it takes seconds to conclusively say whether that hash was previously in the environment. When coupled with hash aliasing, this is a game changer in how security teams operate and the certainty with which they can communicate those results.

Because Chronicle is cloud native, it is trivial to bring in full telemetry data from other cloud solutions, including IaaS and PaaS deployments in public cloud platforms. But most EDR is also running in the cloud, so the data you're already sending to the cloud is easily replicated to Chronicle in most cases without any additional costs to a customer (see Figure 3).

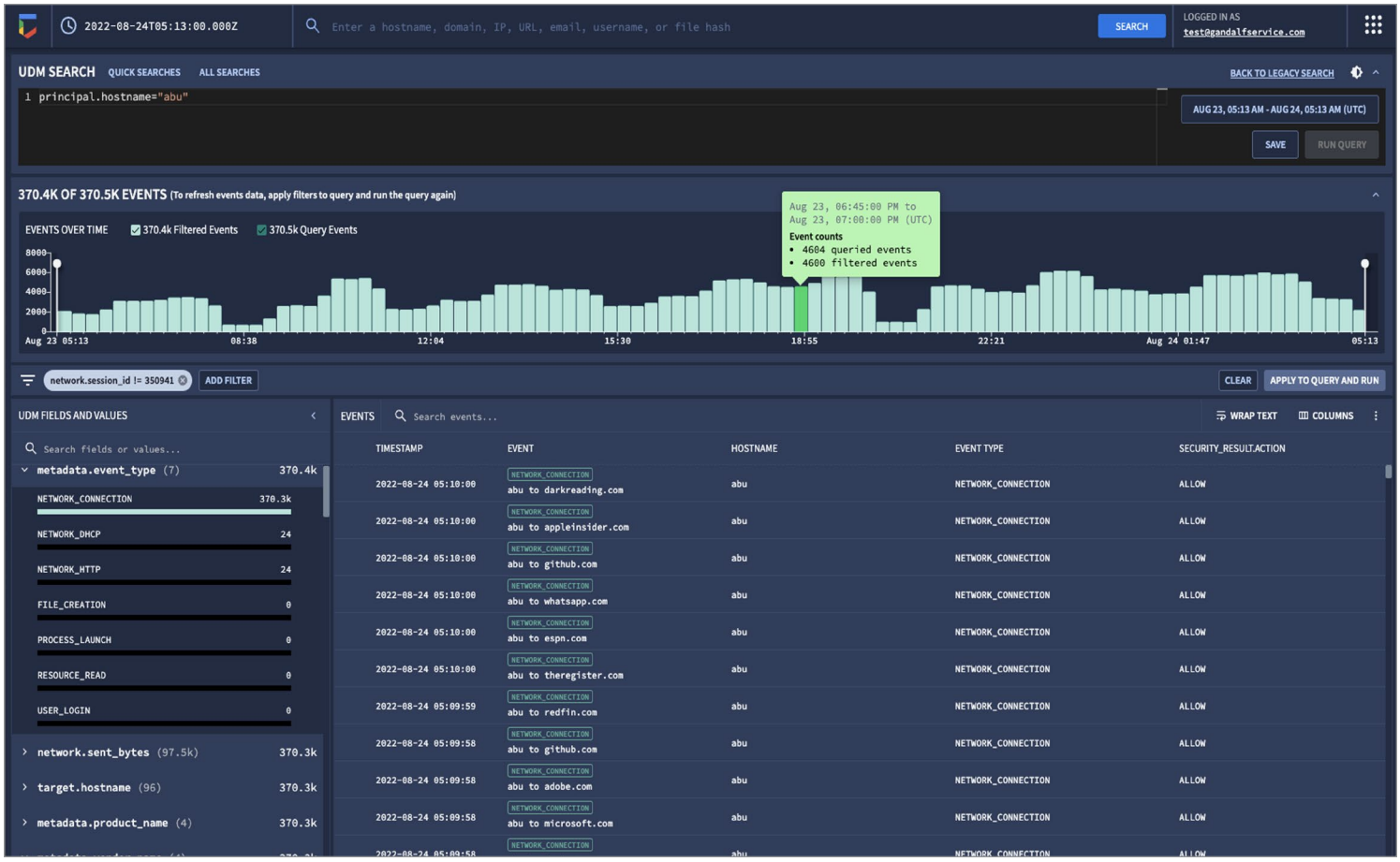

Figure 3. Streaming Data View on the UDM Search Console for Further Analysis

Because Chronicle logs all data, it is trivial to build process trees to provide the analyst context about what is executing on their systems. Most analysts dealing with investigations have been hampered by the inability to easily pivot from an event of interest through child and parent processes. Chronicle not only makes this pivoting fast, but it also supports the notion of process trees.

Modern Threat Detection

As threat actors continue to innovate, they become harder to catch. More precisely, it requires more data to uncover the activities of these actors. "Just log everything" sounds great in theory, but traditionally has been a problem for organizations. This is particularly true at scale. Separately, knowing what data to acquire is often a problem, particularly for smaller organizations and those without dedicated detection engineering teams.

Every time there's another security conference (or seemingly, long weekend), someone drops a new exploit, post-exploitation technique, or security control bypass. This leaves teams racing to understand the issue and the telemetry required to detect it. That, of course, is a precondition for writing the detection logic and testing it in the SIEM. This activity, like so many others involving security monitoring, disproportionately impacts smaller security teams.

Organizations need modern threat detection that operates at scale, and Chronicle delivers. In our tests, we observed multiple threat detections that most organizations haven't even pondered, including in their SIEM. These are enabled by Google's team of detection engineers (see Figure 4), the amount of data ingested into the SIEM, and the ability to perform detections on enriched data in addition to source data. Of course, if your organization does have detection engineering in-house, Chronicle supports the creation of custom detections. It's just a question of how much you need to do rather than letting someone else handle it for you.

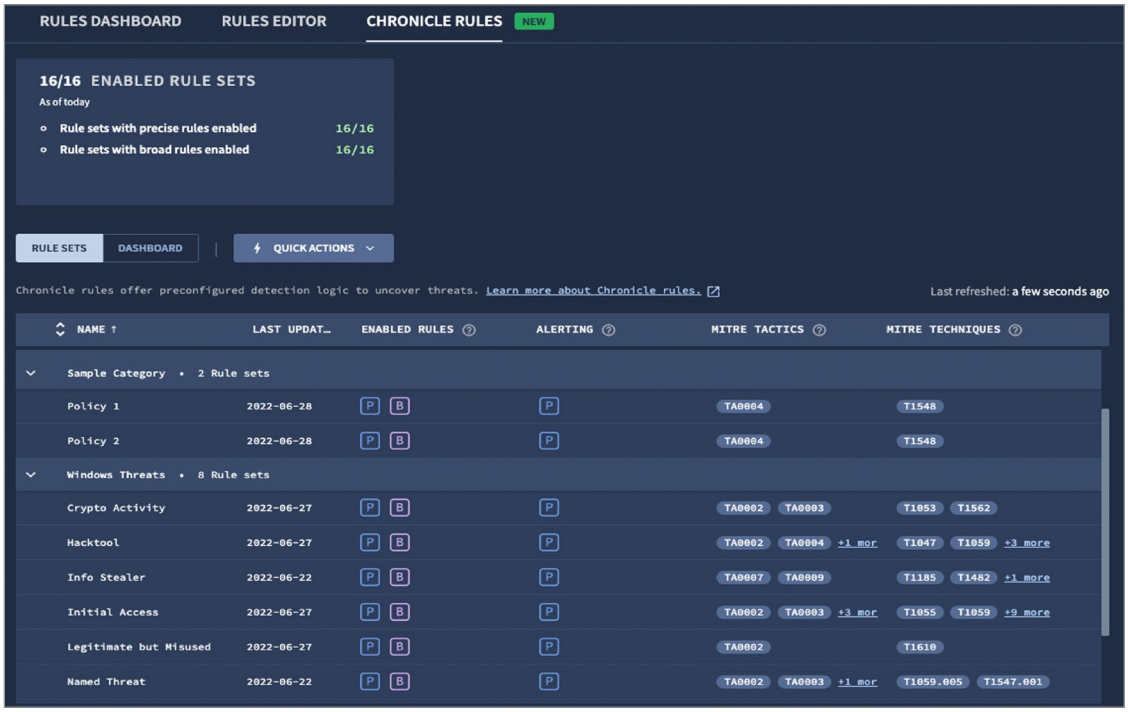

Figure 4. Google's Out-of-the-Box Curated Detections

Detection Engineering

We've mentioned detection engineering a few times already, but it's worth mentioning that Chronicle includes a dedicated detection authoring platform. The detection engineering platform includes the ability to use regular expressions on any field of data recognized by the Chronicle platform (see Figure 5).

```
meta:
  author = "Chronicle Security"
  description = "Detects lolbin not in the expected default directory, i.e., actor has copied it to alternate path to avo
  severity = "HIGH"

events:
  $e.metadata.event_type = "PROCESS_LAUNCH"

  // for consistency, use lowercase, and remove the domain name
  $host = strings.to_lower(re.capture($e.principal.hostname,`^(.*?)(?:\.|$)`))

  // restrict to Microsoft Event Logs & Microsoft Sysmon
  // - change for your EDR as needed
  $e.metadata.vendor_name = "Microsoft"

  // extract the path from principal, excluding volume or unc
  // - re.capture uses string constants so backslashes must be escaped
  $path = re.capture($e.target.process.file.full_path, "^.*?\\\\(.*\\\\).*$")

  // extract the process from target
  $process = re.capture($e.target.process.file.full_path, "^.*?\\\\.*\\\\(.*)$")

  // compare the extracted process against our reference list of monitored lolbins
  $process in %loblas_execute

  // regex of known normal default paths
  // - exluding volume or mount path
  (
```

Figure 5. Custom Detections with Chronicle's Detection Authoring Platform

A consistent problem with detections is determining whether a threat actor bypassed the subject of the detection before the rule was in place. Mature organizations use a technique called "retro hunting" to apply a new detection across previously recorded data. Retro hunts are sometimes complicated by the fact that searching through historical data may require different syntax than real-time detections that fire as new data is ingested. Of course, retro hunts are also limited by the amount of data retained and the time required to search through logs in cold storage.

As previously discussed, Chronicle solves the issue of time required to search because it operates at sub-second speed across terabytes of data and there is no notion of cold storage. Additionally, YARA-L used by Chronicle is also an enabler because there's no notion of differentiating real-time queries from retro hunts. This means that retro hunts can be performed across all data for any new detections you've created with just a single rule, ensuring no threat actor behavior present in your data remains undiscovered.

Chronicle also includes the notion of risk scores that can be added to detections. This directly supports automation with SOAR (more on this later), offering a way to bake priorities into detections themselves. This keeps risk prioritization closest to the tooling where the risk itself is measured, ensuring consistent delivery of security. As any detection engineer will tell you, the schema for data storage in the SIEM is a huge limiting factor. We often analogize this to a set of building blocks. Although you can stack and assemble the blocks in any manner you like, you can't build something requiring components you don't have. To that point, Chronicle supports assigning any value to an array, so detection engineers aren't arbitrarily limited. Consider, for instance, something like a DNS reply. A DNS reply might only return a single IP address. But how should the SIEM handle a reply that returns several IP addresses? Some SIEMs only store the first IP returned; others store only the last. Still others concatenate the results together and store them in a free text field (and we've already discussed why this can be an issue). But because Chronicle supports storing (and searching) data in lists, your flexibility isn't sacrificed in creating detections.

In addition to DNS replies, other areas where lists are extremely useful might include:

- TLS cipher suites supported
- Child process names
- Files downloaded

Enriching data provides context around alarms. As enriched data is stored in Chronicle, context-aware detections can be created to alert from any number of other sources. Chronicle integrates with multiple sources out of the box for context-aware detections, including:

- Asset management databases
- Configuration management systems
- Vulnerability scan data

The power of context-aware detections is hard to overstate. It allows engineers to build rules completely in the SIEM that require SOAR (or other automation tools) to accomplish in other platforms.

Detections can be built using a custom language that looks a lot like YARA. It includes multiple features useful for detections, including variables and the ability to include references from other detections. The ability to create references is another feature missing in most SIEMs and is a game changer when updates to existing detections are required, such as for tuning out a false positive detection. When references are supported, the analyst simply makes a change in one location, and those changes automatically propagate to all detections that use it. This ensures more consistent security delivery and minimizes total cost of ownership. Things that may seem like small details make for huge changes in outcomes.

Event Sequencing

It's hard to believe we saved this amazing feature for the end of our review of the Chronicle platform, but it was necessary because it builds on so many of the other awesome features already discussed. Many detection rules in today's SIEMs include threshold features, such as "10 failed logins from a given IP in five minutes." But rarely do SIEM platforms include robust logic to examine the timestamps in different events to support building detection rules. Chronicle supports this extremely useful feature, something it appropriately refers to as "event sequencing."

Event sequencing can be used to examine a string of events together, in a specified order, within a specific time threshold, to generate a detection. But it goes so much further than this. By supporting detection logic to interrogate data (including enriched data), detection rules can become far more complex. Consider, for instance, the ability to create a rule detecting a successful logon to one account immediately after a failed logon to a completely different account (see Figure 6).

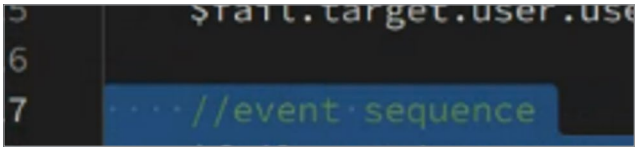

Figure 6. Event Sequencing Logic

There are myriad use cases for event sequencing in creating extremely robust detections. Event sequencing is also useful in reducing false positive detections, enabling the creation of rules with this enhanced logic that would otherwise drown teams in errant alarms.

Response Automation

Chronicle supports integration with SOAR platforms, including Chronicle SOAR. Most SOAR playbooks are used to perform automated actions in response to an alarm detected by some external system, including, of course, SIEM. Common automation actions include examples such as:

- Locking out a compromised account
- Adding a firewall rule to block an IP address on the firewall
- Purging all emails from a sender identified as potentially malicious

The two largest challenges in automating security response are integrating with a detection source and reducing false positives. Chronicle handles the first of these challenges because it supports seamless integration with existing major SOAR platforms and offers API integrations for maximum flexibility. But because this is table stakes for most SIEMs, let's dive deeper into the false positive reduction.

False positive detections are undesirable in any case, but when combined with SOAR, their impact is intensified. Without automation, an analyst will examine the alarm and determine what (if any) action to take. If the alarm is a false positive, there's still an expectation that an analyst will be in the loop to take the appropriate action. But when it comes to SOAR, most playbooks are executed without additional human interaction. False positives typically result in some denial of service condition when coupled with SOAR. Therefore, reducing false positives is critical with automation.

Chronicle supports the reduction of automation-induced denial of service in two primary ways. First, because Chronicle operates on more data and higher quality enrichment, it generates fewer false positives in the first place. Second, in cases where false positives cannot be completely eliminated before sending an alarm to the SOAR platform (a rare occasion given the features in Chronicle), the SOAR platform can more easily diagnose a false positive as part of the playbook logic. Simply put, the more data the SOAR platform is passed from the detection engine, the easier it is to identify an errant detection. With Chronicle SOAR, Google now has capabilities to automatically group related alerts into threat-centric cases and provide seamless automation and response capabilities.

Next Steps

We encourage any security team investigating modernizing their security detections to examine the Chronicle SIEM product. In our evaluation, we found it to be a paradigm changer in how security investigations are conducted and believe it will be a force multiplier for most security teams.

Sponsor

SANS would like to thank this paper's sponsor: